

1 Análise espectral de séries temporais representativas de processos reais e sua filtragem

1.1 Introdução

Neste projeto vamos proceder à análise de séries de dados onde se encontram sobrepostos processos com diferentes escalas. Em cada turma prática, cada grupo fará a análise de um conjunto de dados diferente e deverá desenvolver procedimentos adaptados aos dados que vai analisar. Dependendo dos dados, poderá ser necessário isolar oscilações “rápidas” ou “lentas”, por filtragem, e/ou localizar oscilações transientes, i.e. representativas de processos que só ocorrem num período localizado. O impacto da filtragem será monitorizado com o cálculo de espectros dos dados em diferentes fases do processamento.

1.2 Descrição sucinta das séries de dados

Cada grupo analisará unicamente uma das séries fornecidas, consoante o seu número de ordem, sucintamente descritas nas sub-secções seguintes.

1.2.1 Dados de pressão no fundo do oceano na região do Alaska (Alaska526N1569W4512Z.zip)

Dados contidos no ficheiro Alaska526N1569W4512Z.dat, disponibilizados pela NOAA (NOAA tsunami database), contém 1 ano (2006) de observações de pressão (em psia) num ponto na região do Alaska (52.6N 156.9W) a 4512m de profundidade, com uma amostragem de 15s. Dados em ascii: [Ano Mês Dia Hora Min Sec Pressão]. O sinal predominante nesses dados é a maré. No dia 14 de Novembro de 2006 teve lugar um sismo de grande magnitude nas ilhas Kuril (Curilhas), [46°31'N 153°19'E](#) [46.52°N 153.32°E](#) [Coordinates](#): às 11:14:16 UTC, que deu origem a um pequeno tsunami.

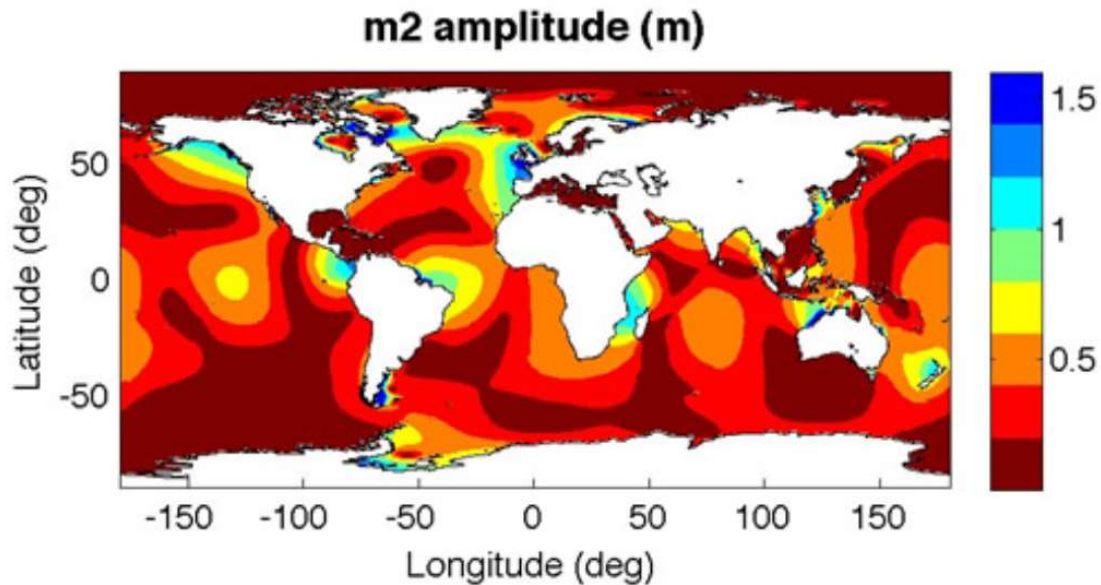


Figura 1-1 – Amplitude da componente dominante da maré (fonte Universidade de Washington, <http://faculty.washington.edu/pmacc/tools/tide.html>)

Sugestões: caracterizar a maré (amplitude dos termos dominantes (lunar e solar), e sua interação (marés vivas e mortas); cf Figura 1-1; filtrar o sinal eliminando as baixas frequências i.e. períodos acima de poucas h; analisar o sinal antes e depois de filtrado no dia 14 de Novembro; caracterizar o tsunami; estimar o tempo de propagação do tsunami.

Notas. Em boa aproximação o tsunami propaga-se à velocidade $c = \sqrt{gH}$, onde H é a profundidade. As ondas de tsunami têm períodos típicos de vários minutos, sendo muito bem amostradas por este sistema. No entanto, sinais sísmicos podem também ser registados com uma amostragem mais ou menos deficiente (dependendo do sinal). A velocidade de propagação pode ser utilizada para identificar o tsunami. É necessário converter psia (pounds per square inch) em m de água equivalentes.

1.2.2 Número diário de manchas solares (isunspots_daily.zip)

O número de manchas solares é um parâmetro muito relevante da variabilidade solar, inicialmente descoberto por Galileo. Na série fornecida:

(https://climexp.knmi.nl/getindices.cgi?WMO=SIDCData/sunspots_daily&STATION=sunspots&TYPE=i&id=someone@somewhere&NPERYEAR=366)

podem identificar-se diversas oscilações, em especial as devidas à rotação solar (27 dias) e às tempestades magnéticas solares (11 anos). Ver o artigo

<https://www.aanda.org/articles/aa/pdf/2006/23/aa4978-06.pdf>.

1.2.3 Anomalia da duração do dia sideral (ilod.zip)

O dia sideral (cf. Diferente do dia solar) oscila devido à troca de momento angular entre a litosfera e outras componente do sistema terra (atmosfera, oceano, terra profunda).

Dados:

https://climexp.knmi.nl/getindices.cgi?WMO=IERSData/lod&STATION=lenth_of_day&TYPE=i&id=someone@somewhere&NPERYEAR=366

1.2.4 Sismo em Sumatra (2012.04.11) observado no Alaska (sismo1_AK.zip)

Dados contidos no ficheiro (sismo1AK.txt): 2h de dados sísmicos (correspondentes à velocidade vertical da superfície) amostrados a 50Hz. Nestes dados existe um sinal de baixa frequência ($f < 0.1$ Hz) contendo as ondas sísmicas produzidas num sismo longínquo (em Sumatra) e dados mais localizados em mais alta frequência ($f > 5$ Hz) correspondentes a um sismo local, efeito secundário do sismo longínquo.

Sugestões: caracterizar o sinal do sismo longínquo (frequências dominantes e amplitudes nas 3 componentes) e do sismo local; em cada caso escolher uma janela que contenha o sinal correspondente e fazer a sua análise espectral; no caso do sismo local sobreposto filtrar o sinal (passa-alto) para isolar o sismo local e proceder de novo à análise espectral; no caso do sismo longínquo calcular a aceleração e o deslocamento e os correspondentes espectros e compará-los com o espectro das velocidades. Discutir.

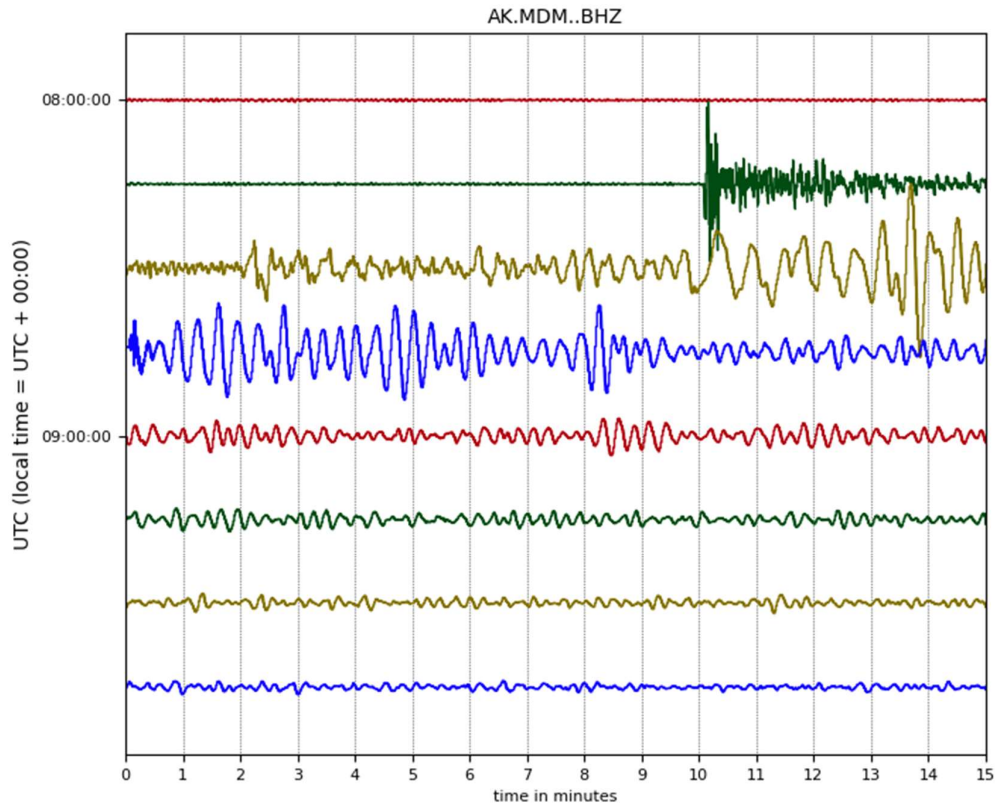


Figura 1-2 – Velocidade vertical no dia 2011.04.11, amostrada a 50Hz. Neste sinal “só” se vêem as ondas de superfície (“longo” período). Notar o sismo local cerca das 9:20 (curto período, curva verde).

1.2.5 Ruído oceânico observado por sismómetro em Portugal em dia de intensa ondulação (sismo1_PT.zip)

Dados contidos no ficheiro “sismo1_PT.dat”: 2 h de dados sísmicos (correspondentes à velocidade vertical da superfície) amostrados a 50Hz (Figura 1-3). Dados em ascii: [Sec Ux Uy Uz]. Nestes dados existe um sinal de baixa frequência contendo as ondas sísmicas produzidas agitação marítima e dados mais localizados em mais alta frequência correspondentes a pequenos sismos locais independentes (um deles identificado na **Error! Reference source not found.**, com um sistema avançado de filtragem).

Sugestões: caracterizar o sinal completo (frequências dominantes e amplitudes nas 3 componentes); isolar o sismo local com um filtro passa-alto (e.g. $f > 5\text{Hz}$) por média-móvel e por filtro no espectro; proceder à análise com janelas de 1 minuto.

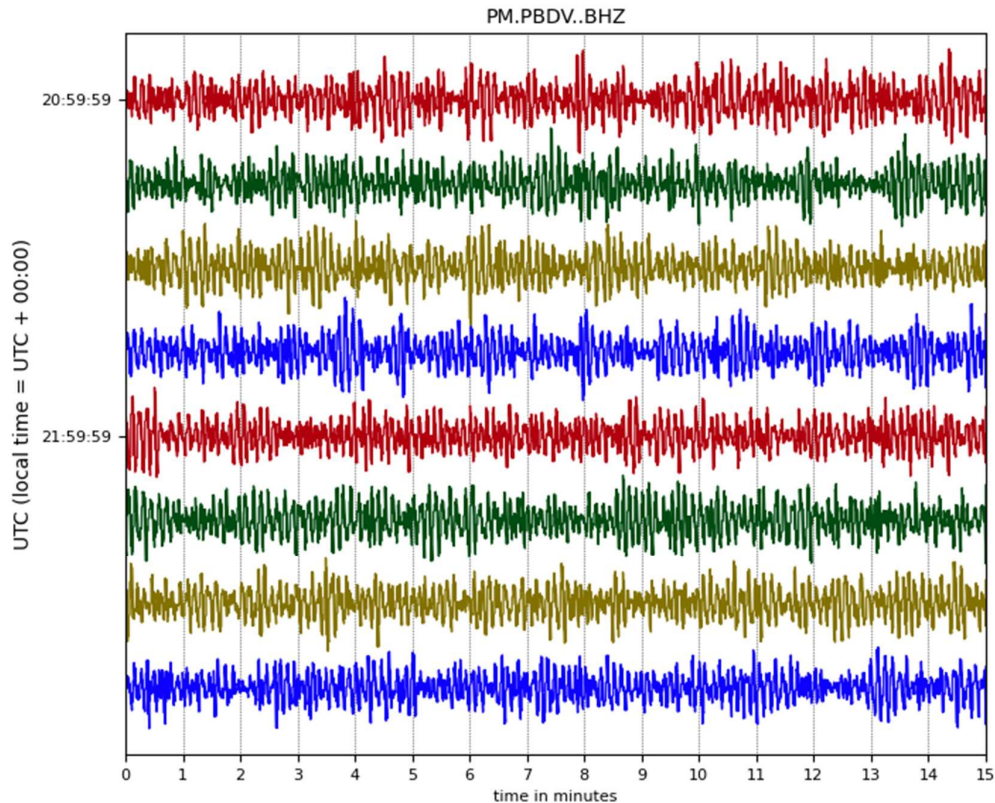


Figura 1-3 – Velocidade vertical no dia 2011.02.17, amostrada a 50Hz. O sinal dominante com 5-20s de período é ruído oceânico. O sinal contém no entanto pequenos sismos locais de mais alta frequência (e.g. 5 a 20 Hz). Um deles está representado na **Error! Reference source not found.**

1.2.6 Radiação solar medida no campus da FCUL (RadFCUL2015.zip)

O ficheiro RadFCUL2005.xlsx contém observações de radiação global, difusa e direta, com amostragem de 1 minuto, no campus da FCUL em Lisboa, durante 1 ano completo (mas não um ano civil). Tais observações contém um forte ciclo diurno e um ciclo anual, variabilidade sinóptica (meteorologia) e de pequena escala, nomeadamente sombreamentos por nuvens. Os dados são afetados no início da manhã e no fim da tarde por sombreamento espúrio (edifícios, árvores).

Este dataset contém medições de radiação solar registadas por uma estação meteorológica instalada no Campus Solar da FCUL. Registrando as componentes global e difusa horizontal, além de directa normal, estes dados são locais e possuem uma resolução temporal de 1 min. O período compreendido por estes dados é de 09/01/2015 a 08/01/2016, o que permite confirmar a existência de um ciclo diário e anual. Além disso, com a aquisição de dados tão frequente, torna-se possível avaliar o impacto da degradação da resolução temporal e/ou do registo de valores médios (e não pontuais) na deteção de variabilidade intra-diária (introduzidos pela meteorologia).

Nota: Como dados reais que são, este dataset é afetado por falhas de aquisição (devidamente identificadas com o valor “-999”) e a um erro de leitura sistemático negativo reduzido mas que, p.e., resulta em valores de radiação negativos durante a noite.

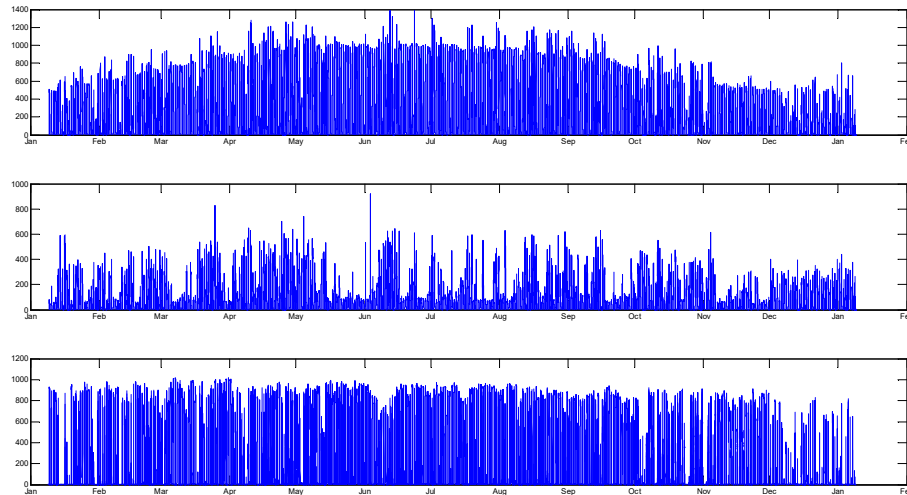


Figura 1-4 – Radiação Solar em Lisboa (2015-2016), FCUL. Fonte IDL (Rodrigo Silva e Miguel Brito)

Sugestões: limpar a série de erros e interpolar para preencher os buracos; estimar a amplitude dos ciclos diurno e anual em cada uma das séries de dados; filtrar o ciclo diurno e anual e analisar a série filtrada (cf. Método); Analisar as séries filtrada nas altas frequências (períodos sub-horários) e na escala sinóptica (3-10 dias). Calcular a série anual de radiação global diária (1 valor por dia) e recalculer o seu ciclo anual.

1.2.7 Evolução da temperatura e da concentração atmosférica de CO_2 em bolhas de gelo em Vostok (Vostok.zip)

Os dados a analisar incluem estimativas da evolução da temperatura Terra (Figura 1-5) nos últimos 420000 anos e da concentração atmosférica de CO_2 , obtidas por análise de bolhas de ar retidas no gelo em Vostok (Antárctica). Dada a natureza do processo de análise, as amostras têm resolução temporal variável (mais fina nos dados recentes, mais grosseira nos dados antigos) e não existe sincronização entre as duas séries. No entanto, ambas as séries apresentam variabilidade intensa com períodos associados a oscilações de Milankovic (e.g. cerca de 100ky, 40ky e 20ky) e oscilações mais rápidas.

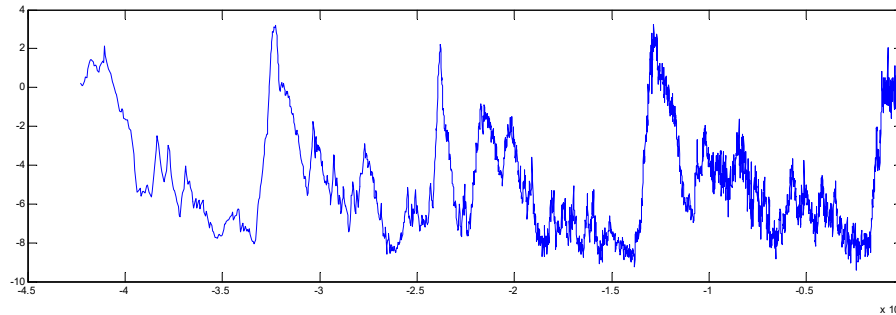


Figura 1-5 – Temperatura nos últimos 420 ky, estimada a por análise de bolhas de ar no gelo em Vostok (Antártica). Fonte Jouzel et al. (Nature, 329:403-8), Petit et al. (Nature, 399: 429-436).

Sugestões: ler ambas as séries e interpolá-las para a mesma base temporal com resolução comparável com a mais alta resolução disponível; calcular o seu espectro de amplitude, estimando a amplitude das oscilações dominantes; verificar o resultado por observação visual dos gráficos; filtrar os dados de forma a eliminar oscilações com período superiores a 10000 anos; analisar o espectro de amplitude das séries resultantes; estudar a correlação entre as duas séries.

1.2.8 Evolução de variáveis num cenário de aquecimento global (RACMO38_7.zip)

Os dados aqui apresentados foram obtidos no projeto EURO-Cordex e são constituídos por uma série temporal de valores diários da temperatura máxima, mínima, precipitação (mm/s) e humidade específica, num cenário de aquecimento global RCP8.5 (2006-2100).

Sugestões: caracterizar ciclo diurno e anual das diferentes variáveis; filtrar; voltar a caracterizar o sinal filtrado (espectro); Identificar tendências. Notar que estes dados incluem tendências acentuadas. Sugere-se que a caracterização seja feita para janelas corridas de 30 anos (2011-2030...2071-2100).

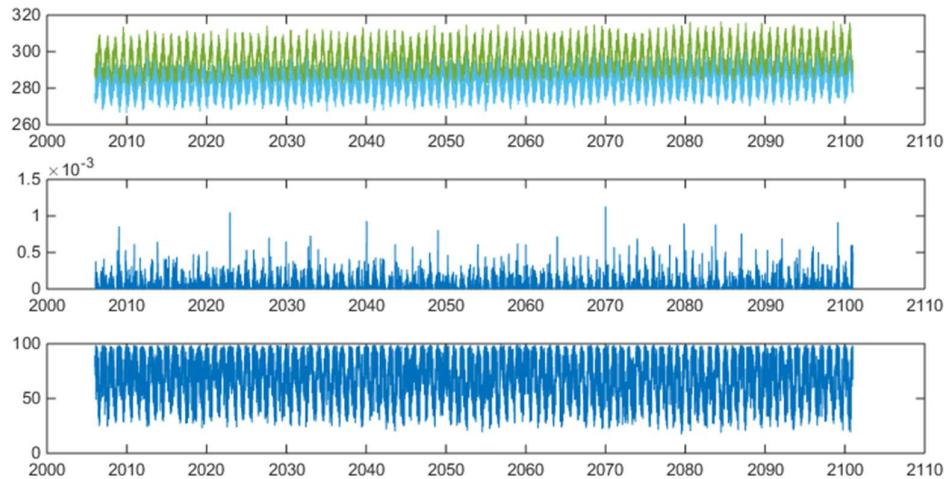


Figura 1-6 – Temperatura aos 2m (min e max), precipitação e humidade relativa no período 2006-2100, com amostragem diária, no cenário RCP8.5 a 38N, 7W, simulado pelo modelo RACMO (EURO-Cordex). Circulação atlântica meridional AMOC (imoc_mar_hc10_day.zip)

A circulação meridional Atlântica exhibe oscilações à escala sazonal, interanual e decadal. A sua variabilidade é relevante da variabilidade lenta da região.

https://climexp.knmi.nl/getindices.cgi?WMO=NOCDData/moc_mar_hc10_day&STATION=AMOC_26N&TYPE=i&d=someone@somewhere&NPERYEAR=366

1.3 Método

Neste projeto pretendemos analisar séries temporais em diferentes escalas, contendo um ou mais processos periódicos, processos transientes, tendências, oscilações aperiódicas e ruído. Os procedimentos a adotar devem ser adaptados aos dados em questão. Aqui sugere-se um conjunto de procedimentos genéricos a considerar em todos os casos.

- (1) Começar por inspecionar os dados, representando-os graficamente. Verificar se foram amostrados regularmente e a que passo, identificar eventuais erros ou falhas. Verificar e, se necessário corrigir, as unidades utilizadas.
- (2) Corrigir erros, eliminando-os. Se necessário interpolar os dados para recuperar ou construir uma base regular de amostragem.
- (3) Calcular o espectro de amplitude dos dados. Se for conveniente, pode truncar a dimensão da série a um número favorável para a fft (e.g. na forma $2^k 3^m 5^n$).
- (4) Identificar tendências nos dados, i.e. variações cujo período aparente é superior à duração do sinal. Se existir uma tendência eliminá-la por regressão linear ou por média móvel de longo período.
- (5) Identificar ciclos dominantes nos dados (anual, diurno, etc) e desenhar filtros capazes de os remover.
- (6) Analisar os dados filtrados: inspeção da série e cálculo no novo espectro

- (7) Se for apropriado, calcular o espectro em janelas parcelares, no sentido de identificar oscilações transientes, e.g. um tsunami ou um sismo.

1.3.1 Filtros

Os filtros a utilizar podem ser de vários tipos, mas pelo menos dois devem ser ensaiados: filtros de média móvel de peso constante e variável, filtros impostos diretamente no espectro, recuperando a série por transformada inversa. As funções de transferência dos filtros devem ser apresentadas.

1.3.2 Análise espectral por janelas

Esta análise é interessante para identificar processos transientes que só estarão presentes ocasionalmente ou para caracterizar a evolução temporal do espectro.

1.3.3 Etc

Os projetos são abertos. Tentem pensar sobre os dados e sobre os processos que estão na sua origem.

1.4 Protocolo

- (a) Começar por gerar uma série sintética contendo 3 oscilações com períodos de 24h, 29dias, 1ano, respetivamente (escolha amplitudes no intervalo $[1,10]$). Analise essa série espectralmente. Desenhe um filtro para isolar cada uma das oscilações. Verifique o resultado.
- (b) A série de dados a analisar terá o número do seu grupo. Proceda a análise dessa série.

2 Otimização de parâmetros

Neste trabalho vamos utilizar uma variante da técnica de *simulated annealing* para estimar parâmetros de um modelo de ajuste a dados observados. Serão considerados 5 problemas, cada um com duas variantes correspondendo a dados alternativos mas com a mesma física. A técnica proposta foi descrita nas aulas (cf. Powerpoints disponibilizados), sendo idêntica ao *simulated annealing* com a diferença de não considerar a possibilidade de saltos para valores mais elevados da função de custo. Localmente, no entanto, o método pode saltar sobre barreiras (máximos locais) da função de custo, especialmente enquanto se encontra suficientemente “quente”. O algoritmo de otimização encontra-se em

[GitHub - pjmateus/monte-carlo-annealing: Monte-Carlo search for the minimum of the multidimensional "cost" function](https://github.com/pjmateus/monte-carlo-annealing)

O método é descrito em

Miranda, P. M. A., & Mateus, P. (2022). Improved GNSS water vapor tomography with modified mapping functions. *Geophysical Research Letters*, 49, e2022GL100140. <https://doi.org/10.1029/2022GL100140>

2.1 Objetivos dos trabalhos

Em cada um dos trabalhos pretende-se:

- (1) Criar uma série sintética de observações do problema proposto, com a adição opcional de ruído aleatório (0%, 1%, 5%, 10%)
- (2) Representar graficamente os perfis observado antes de fazer a inversão;
- (3) Definir a função de custo adequada para o problema proposto, que será sempre função de 3 parâmetros a otimizar;
- (4) Escolher os limites de variação adequados para os parâmetros e as condições de paragem;
Importante: se uma das soluções ficar muito perto do limite ajustar os limites.
- (5) Adaptar o código no sentido de realizar a otimização dos parâmetros a partir de uma seleção aleatória do estado inicial (*first guess*), notando que há 3 parâmetros a ajustar;
- (6) Iterar o algoritmo até uma solução para os parâmetros (X,Y,Z);
- (7) Repita (6) 10 vezes com diferente estado inicial. Mostre o histograma das soluções.
- (8) Mostrar a estrutura da função de custo em 3 gráficos contourf correspondentes às planos (x=X,y,z), (x,y=Y,z); (x,y,z=Z);
- (9) Realizar outras ações especificamente pedidas para o trabalho correspondente.

2.1.1 Localização de uma pepita de ouro num levantamento gravimétrico

A lei de Newton da gravitação pode escrever-se:

$$\vec{F} = -G \frac{Mm}{r^2} \vec{u}_r \quad (1)$$

Um gravímetro mede a **anomalia da componente vertical da gravidade**. Admita que fez um levantamento gravimétrico ao longo de uma linha reta com 10 m de comprimento com observações de metro a metro, 20 cm acima da superfície, numa região homogénea e plana onde a massa volúmica da crosta vale $2.6 \times 10^3 \text{ kg m}^{-3}$. Nessa região encontra-se, sob a superfície uma pepita de ouro (massa volúmica $19 \times 10^3 \text{ kg m}^{-3}$) esférica. Pretende-se localizar a pepita (X,Z) e estimar a sua massa.

Repita experiência substituindo o ouro por chumbo.

2.1.2 Localização de uma caverna num levantamento gravimétrico

Idêntico ao caso anterior, mas a anomalia de gravidade é devida a uma caverna. Pretende-se localizar a caverna (X,Z) e estimar o seu volume. A caverna tem menos de 20 m^3 e encontra-se a uma profundidade inferior a 500m.

Repita experiência admitindo que a caverna está cheia de água.

2.1.3 Localização e caracterização de um planeta a partir do campo gravítico

Neste caso os dados indicam a posição (x,y) de uma sonda e o valor da aceleração da gravidade medida e o objetivo é localizar o planeta e a sua massa. Admita que a sonda segue uma órbita elíptica completa, com uma excentricidade de 10%, na qual o planeta ocupa um dos focos, com uma massa no intervalo $[0.5 \times 10^{24}, 10^{24}]$. Admita que a gravidade do planeta é a que seria produzido por uma massa pontual no seu centro de massa. **Repita com o primeiro terço das observações e discuta o resultado.**

2.1.4 Localização de uma baleia a partir dos tempos de chegada do som

Considere 4 estações localizadas no fundo do mar, nos quatro cantos duma região com 50 km de lado (x,y) e 5 km de profundidade. Admita que o som se propaga à velocidade de 1500 m/s. **Repita com as estações alinhadas ao longo da diagonal.**

2.1.5 Localização de um canhão a partir da localização de pontos da trajetória dos seus projéteis.

Considere a trajetória parabólica de um projétil num campo de gravidade constante, sem resistência do ar e em rotação. Admita que o canhão se encontra em $z=0$. Considere 10 observações. Estime a posição do canhão (x_0) e a velocidade inicial do projétil ($u; w_0$).

2.1.6 Localização de um canhão a partir da localização de pontos da trajetória dos seus projeteis. (versão 2)

Idêntico ao anterior mas fixando $x=0$. Calcule z_0, u, w_0 .

2.1.7 Localização de uma pepita de ouro num levantamento gravimétrico (versão 2)

Tal como 2.1.1 mas fixando $m=10$ kg, e calculando (X,Y,Z) . As observações serão distribuídas no plano $(X,Y, z=0.2\text{m})$ com 10000 observações.

2.1.8 Localização e caracterização de um planeta a partir do campo gravítico (versão 2)

Neste caso considere uma órbita com excentricidade 0.3.